INFORMATION

Grundbegriffe

Offensichtlich hat Informatik mit Information zu tun. Wir sprechen auch von Informationsverarbeitung, von Informationstechnologie*) und von Informationsgesellschaft. Was aber verstehen wir unter Information?

Unter Information versteht man die Bedeutung, die durch eine Nachricht übermittelt wird. Nachrichten können gesprochene oder geschriebene Texte, Bilder, Geräusche oder sonstige Daten sein, die wir mit unseren Sinnesorganen wahrnehmen und mit technischen Hilfsmitteln speichern und übertragen können. Welche Information einer solchen Nachricht entnommen wird, ist vom Empfänger dieser Nachricht abhängig und somit subjektiv. So kann zum Beispiel ein einfaches Wort (zB "mayday") für den einen Empfänger bedeutungslos sein, für den anderen hingegen eine wichtige Information beinhalten. (Das Wort "mayday" wird als Notsignal im internationalen Funkverkehr verwendet).

Um Nachrichten oder Daten**) speichern und übertragen zu können, benötigen wir Signale. Ein Signal wird durch den Werteverlauf eines physikalischen Mediums repräsentiert. So kann zum Beispiel die Nachricht "mayday" im Morsecode durch eine elektrische Impulsfolge übertragen oder eine gesprochene Nachricht magnetisch auf einer Tonbandkasette aufgezeichnet werden. Im Fall der Impulsfolge handelt es sich um ein digitales Signal, das nur eine bestimmte Anzahl von Werten annehmen kann (zB Strom oder kein Strom). Die magnetische Aufzeichnung auf der Tonbandkasette entspricht ist ein analoges Signal. Analoge Signale haben einen kontinuierlichen Werteverlauf.

Der Unterschied zwischen digitalen und analogen Signalen lässt sich sehr anschaulich am Beispiel von Digital- und Analoguhren illustrieren. Bei einer Analoguhr wird die Zeit (die Nachricht) durch den Winkel des Zeigers dargestellt. Der Zeiger bewegt sich kontinuierlich, es sind somit beliebige Zeitwerte darstellbar. Entsprechend kann auch die Uhrzeit (zumindest prinzipiell) beliebig genau abgelesen werden. Eine Digitaluhr hingegen kann die Zeit nur auf eine bestimmte Ziffernanzahl (zB auf Minuten genau) darstellen. Es sind somit digital nur bestimmte Zeitwerte darstellbar. Bei der Analoguhr ist die Zeigerstellung das Signal, bei der Digitaluhr ist es zB das Leuchten der Ziffern.

---

*) Häufig wird auch von Informations- und Kommunikationstechnik (engl. Information and Communication Technology abgekürzt ICT) gesprochen.

**) Die Begriffe Nachricht und Daten sind hier gleichbedeutend. In der Informationstechnik wird häufig der Begriff Daten (engl. data), in der Kommunikationstechnik eher der Begriff Nachricht (engl. message) verwendet.

Die Nachricht ist die dargestellte Zeitangabe. Der Betrachter der Uhr wieder kann aus dieser Zeitangabe zB die Information entnehmen, dass es schon spät ist oder dass vielleicht die Uhr falsch geht.

Nachrichten werden häufig nach festgelegten Regeln dargestellt. Die Uhrzeit "Viertel vor Mitternacht" kann zB in der Form 23.45 durch eine zweistellige Stundenangabe und eine zweistellige Minutenangabe durch einen Punkt getrennt dargestellt werden. Die äussere Form - zwei Ziffern, ein Punkt und wieder zwei Ziffern - wird als Syntax dieser Darstellung bezeichnet. Entsprechend dieser Syntax ist zB 8.45 eine fehlerhafte Uhrzeit (vor dem Punkt ist nur eine Ziffer, es müsste 08.45 heissen). Ebenso ist 12:00 syntaktisch falsch, weil entsprechend der obigen Syntax zwischen den Ziffern kein Doppelpunkt erlaubt ist. Was erlaubt ist und was nicht erlaubt ist wird von uns willkürlich durch die Syntax festgelegt. Selbstverständlich können beide Schreibweisen in einer geänderten Syntax zulässig sein.

Die Syntax beschreibt nur die äussere Form einer Darstellung, ihre Bedeutung wird durch die Semantik festgelegt. Die Semantik der Darstellung der Uhrzeit im obigen Beispiel kann etwa die ersten beiden Ziffern als Stundenzahl und die letzten beiden Ziffern als Minutenzahl definieren. Die Darstellung 08.75 wäre nach dieser Definition zwar syntaktisch richtig, aber semantisch falsch (falls wir festgelegt haben, dass der Wert der Minutenzahl kleiner als 60 sein muss).

Syntax und Semantik der Darstellung einer Uhrzeit bilden in unserem Beispiel eine künstliche Sprache. Auch natürliche Sprachen (wie Deutsch oder Englisch) haben eine Syntax (bei natürlichen Sprachen als Grammatik bezeichnet) und eine Semantik. So ist zB der Satz "Gold hat im Morgenstund Mund" zwar lexikographisch sortiert, entspricht aber nicht der deutschen Grammatik, der Satz "Der Heuschreck integriert die Zuckerfabrik" hingegen zwar grammatikalisch richtig aber semantischer Unsinn.

Während natürliche Sprachen evolutionär entstanden sind, sind künstliche Sprachen zumeist gezielt für einen bestimmten eingeschränkten Verwendungszweck entwickelt worden. Beispiele für künstliche Sprachen sind etwa

die Formelsprache der Mathematik,
die chemische Formelsprache zur Beschreibung eines Molekülaufbaus,
die Sprache zur Beschreibung von Schachzügen,
die Notenschrift oder
Programmiersprachen.

Künstliche Sprachen haben gegenüber natürlichen Sprachen oft den Vorteil grösserer Prägnanz und Eindeutigkeit der Darstellung.

Die Menge der Zeichen, aus denen Texte einer Sprache gebildet werden können, wird als Alphabet bezeichnet. Das Alphabet der deutschen Sprache zum Beispiel enthält die Gross- und Kleinbuchstaben, aber auch Umlaute, den Zwischenraum und Interpunktionszeichen.*)

Weitere Beispiele sind etwa das Alphabet der 24 griechischen Kleinbuchstaben:

{α|β|γ|δ|ε|ζ|η|ϑ|ι|κ|λ|μ|ν|ξ|ο|π|ρ|σ|τ|υ|φ|χ|ψ|ω}

das Alphabet der vier Spielkarten:

{♣|♦|♥|♠}

das Alphabet der 10 Dezimalziffern:

{0|1|2|3|4|5|6|7|8|9}

oder das Alphabet der 12 Tierkreiszeichen:

{♈|♉|♊|♋|♌|♍|♎|♏|♐|♑|♒|♓}

Die Zeichen { , } und | sind nicht Bestandteil des jeweiligen Alphabets, sondern dienen vielmehr dazu, den Anfang und das Ende der Zeichenfolge zu kennzeichnen beziehungsweise als Trennzeichen zwischen den einzelnen Zeichen des Alphabets. Genau genommen wird versucht, eine Sprache mit Hilfe einer anderen Sprache zu beschreiben. Diese übergeordnete, beschreibende Sprache bezeichnen wir als Metasprache. Die Zeichen { , } und | bilden daher das Alphabet dieser Metasprache.

Ein Alphabet, das nur aus zwei Zeichen besteht, nennt man Binäralphabet und die beiden Zeichen Binärzeichen.

Beispiele für Binäralphabete sind

{●|❍}
{+|–}
{❍|×}
{Ø|1}

*) Die Schriftzeichen haben ihren Ursprung in einer in Mesopotamien entstandenen Bilderschrift, deren Alphabet aus mehreren tausend Zeichen bestand. Im dritten Jahrtausend vor Christus wurde dieses Alphabet auf die 560 Zeichen der Keilschrift reduziert. Die chinesische Schrift kennt heute nochüber 40000 Einzelzeichen von denen jedoch nach deer Reform nur noch etwa 3000 bis 4000 verwendet werden.

Binäralphabete haben den Vorteil, dass sie sich leicht einfachen Signalen zuordnen lassen. So können zB die physikalischen Zustände

hell - dunkel
offen - geschlossen
positive Ladung - negative Ladung
hohe Spannung - keine Spannung

für die Darstellung der beiden Binärzeichen verwendet werden.

Codierung

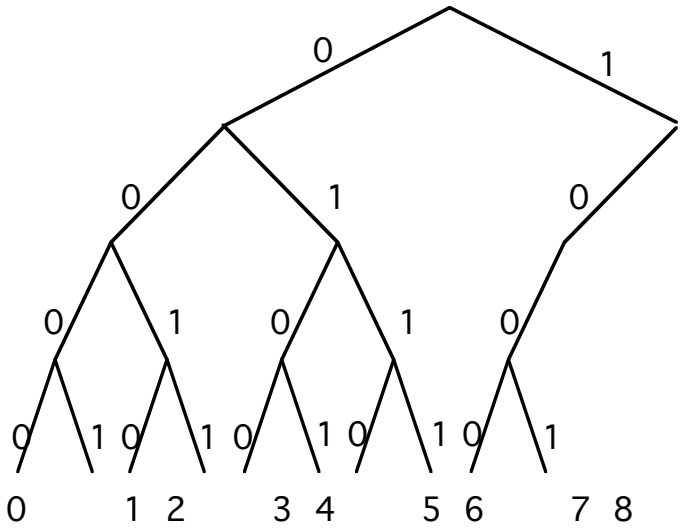
Um diesen Vorteil der technisch einfachen Darstellbarkeit auch für Alphabete nutzen zu können, die mehr als zwei Zeichen umfassen, kann es zweckdienlich sein, die Zeichen beliebiger Alphabete auf binäre Zeichen abzubilden. Ein Beispiel für eine solche Abbildung ist die Darstellung der Dezimalziffern durch einen Binärcode:

0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

Jeder Dezimalziffer wird durch die obige eindeutig Tabelle ein Folge von vier binären Zeichen zugeordnet (für die binären Zeichen werden hier die Symbole 0 und 1 verwendet). Diese Zuordnung wird als Binärcode bezeichnet. Die Folge von binären Zeichen, die einer Dezimalziffer entspricht ist ein Codewort. (Die Analogie zu den Wörtern natürlicher Sprachen, die ja auch durch eine Folge von Buchstaben dargestellt werden ist offensichtlich). Während in natürlichen Sprachen die Wörter variable Länge haben und daher durch Trennzeichen (zB einen Zwischenraum) voneinander getrennt werden müssen, haben die Wörter unseres Binärcodes feste Länge. Trennzeichen werden dadurch überflüssig.

Die binären Zeichen 0 und 1 werden auch als Binärziffern beziehungsweise als Bit (Abkürzung für binary digit) bezeichnet.

Selbstverständlich können die Dezimalziffern auch durch einen völlig anderen als den oben gezeigten Binärcode dargestellt werden. Der hier verwendete Code besticht jedoch durch eine hohe Systematik, die bei einer graphischen Darstellung des Codes durch den sogenannten Codebaum deutlich wird:



Der Codebaum enthält ebenso wie die Codetabelle die gesamte Information, die zur Codierung (das ist die Umwandlung einer gegebenen Dezimalziffer in das entsprechende binäre Wort) sowie auch zur Decodierung (das ist die Rückwandlung eines gegebenen binären Wortes in die zugehörige Dezimalziffer) benötigt wird.

Aufgrund der Systematik des verwendeten Codes können Codierung und Decodierung aber auch durch eine mathematische Formel beschrieben werden:

$$w = \sum z_i 2^i$$

wobei w der Wert der Dezimalziffer, z_i der Wert der i-ten Binärziffer und i der Index (das ist die Nummer) der jeweiligen Binärziffer ist, wenn diese von rechts nach links und mit Null beginnend durchnumeriert werden.

Lautet zB der Binärcode 0101 so ist der dezimale Wert

$$w = 0*2^3+1*2^2+0*2^1+1*2^0 = 0*8+1*4+0*2+1*1 = 4+1 = 5$$

Ein einfacheres Verfahren zur Berechnung des Wertes einer binärcodierten Dezimalzahl (Decodierung) verläuft wie folgt:

> Man nehme den Wert der vordersten Binärziffer, verdopple diesen und addiere dazu die nächste Binärziffer, verdopple diese Summe, addiere die nächste Binärziffer und so fort bis die letzte Binärziffer addiert worden ist. Das Ergebnis ist der Wert der Dezimalzahl.

Im Beispiel 0101 ergibt sich der dezimale Wert w zu

$$w = ((0*2+1)*2+0)*2+1 = (1*2+0)*2+1 = 2*2+1= 4+1 = 5$$

Ebenso können aus dem Wert der Dezimalzahl die Binärziffern (von rechts nach links) ermittelt werden. Die Codierung verläuft wie folgt:

> Ist der Wert der Dezimalzahl gerade, so ist die letzte Binärziffer 0, andernfalls 1. Nun wird die Dezimalzahl ganzzahlig halbiert. Ist das Ergebnis gerade, so ist die nächste Binärziffer 0, andernfalls 1. Danach wird wieder ganzzahlig halbiert und so fort. Das Verfahren wird beendet, wenn die gewünschte Anzahl von Binärziffern ermittelt ist.

Für die Dezimalzahl 5 erhält man zB

> 5 ist ungerade, daher ist die letzte Binärziffer 1
> 5÷2 = 2 ist gerade, daher ist die nächste Binärziffer 0
> 2÷2 = 1 ist ungerade, daher ist die nächste Binärziffer 1
> 1÷2 = 0 ist gerade, daher ist die nächste Binärziffer 0

Von rechts nach links angeschrieben ergeben die Binärziffern das Wort 0101.

Die Verfahren zur Codierung und zu Decodierung sind Beispiele für Algorithmen (diese werden später ausführlich besprochen). Codetabellen, Codebäume, Formeln und Algorithmen sind somit unterschiedliche Beschreibungs- oder Notationsformen für einen Code. Die Informatik beschäftigt sich ausführlich mit solchen Notationsformen.

Beide Verfahren sind auch für mehrstellige Dezimalzahlen anwendbar. Durch einen binären Code der Länge l lassen sich

$$n = 2^l$$

unterschiedliche Worte bilden. Dadurch lassen sich positive ganze Dezimalzahlen zwischen Null und n-1 darstellen.

Durch einen Binärcode der Länge 4 können somit die 16 Hexadezimalziffern von 0 bis 15 dargestellt werden (Hexadezimalziffern sind die Ziffern des hexadezimalen Zahlensystems mit der Basis 16).

Binärcodes der Länge 8 erlauben die Darstellung von 2^8=256 unterschiedlichen Zeichen (8 zusammengehörige Bits werden auch als Byte bezeichnet). Ein solcher 8 Bit Code wird zB zur Verschlüsselung von Gross- und Kleinbuchstaben sowie Ziffern und Sonderzeichen im sogenannten ASCII-Code *) verwendet.

In der folgenden Tabelle sind alle druckbaren ASCII-Zeichen in der Reihenfolge des ASCII-Codes dargestellt (die nicht druckbaren Zeichen sind durch das Zeichen □ repräsentiert):

```
 ÒÚÛÙı ˆ˜¯˘˙ ¸˝˛ˇ
□                -
  !"#$%&'()*+,-./
0123456789:;<=>?
@ABCDEFGHIJKLMNO
PQRSTUVWXYZ[\]^_
`abcdefghijklmno
pqrstuvwxyz{|}~□
ÄÅÇÉÑÖÜáàâäãåçéè
êëíìîïñóòôöõúùûü
†°¢£§•¶ß®©™´¨≠ÆØ
∞±≤≥¥µ∂∑∏π∫ªºΩæø
¿¡¬√ƒ≈Δ«»… ÀÃÕŒœ
––“”‘’÷◊ÿŸ⁄€‹›fifl
‡·‚„‰ÂÊÁËÈÍÎÏÌÓÔ
 ÒÚÛÙı ˆ˜¯˘˙ ¸˝˛ˇ
```

Durch einen Code können nicht nur Zahlen sondern beliebige Daten verschlüsselt werden. Alles was hier über Binärcodes gesagt wurde gilt sinngemäss auch für die Verschlüsselung durch Zeichen eines beliebigen Alphabets.

*) ASCII steht für American Standard Code for Information Interchange

Beispiele für solche Codes sind etwa der aus drei Buchstaben gebildete Codes für Währungen zB

CHF, USD, EUR, ...

die Verschlüsselung der Flughäfen, zB

ZHR, VIE, JFK, CDG, ...

oder die internationalen Autokennzeichen, zB

CH, A, D, FL, ...

aber auch die Codes für Kleidergrössen

XS, S, M, L, XL, XXL

für Monatsnamen

JAN, FEB, MRZ, APR, ...

für Himmelsrichtungen

N, NE, E, SE, S, SW, W, NW

für Papierformate, zB

A1, A2, A3, A4, A5

oder für Stimmungslagen, zB

:-), :-(, ;-)

Zusammenfassend lässt sich sagen, dass Daten und Nachrichten entsprechend einer Syntax aus einzelnen Zeichen eines Alphabets zusammengesetzt dargestellt werden. Mittels eines Codes können die Wörter eines Alphabets auf die Wörter eines anderen Alphabets abgebildet werden damit diese letztlich durch physikalische Signale aufgezeichnet oder übertragen werden können. Während diese Signale real existieren und zB beobachtet und gemessen werden können, sind die durch diese Signale repräsentierten Daten und Nachrichten abstrakte Interpretationen dieser Signale.

Die folgende Tabelle gibt Beispiele für Signale und deren Interpretation im Alltag:

Signal	Nachricht
rote Ampel	Stop
erhobener Daumen	Zustimmung
Kirchturmschläge	Zeitangabe
Sirenengeheul	Alarm

Informationstheorie

Welche Information jedoch der Empfänger einer Nachricht aus dieser entnimmt hängt subjektiv vom Empfänger ab. Die von Shannon 1948 entwickelte Informationstheorie versucht dennoch, ein Mass für den Gehalt der durch eine Nachricht übermittelten Information zu definieren.

Dieser Informationsgehalt soll unabhängig von der Form der Codierung dieser Nachricht ausschliesslich von der Wahrscheinlichkeit abhängen, mit der der Empfänger diese Nachricht erwartet. Dabei soll Nachrichten, die mit hoher Wahrscheinlichkeit erwartet werden ein niedriger, solche die mit geringer Wahrscheinlichkeit erwartet werden hingegen ein hoher Informationsgehalt entsprechen.

Dieser Forderung wird entsprochen, wenn man als Informationsgehalt h eine monoton wachsende Funktion f des Reziprokwertes der erwarteten Wahrscheinlichkeit p der Nachricht wählt:

$$h = f(1/p)$$

Betrachtet man den Informationsgehalt einer aus mehreren voneinander unabhängigen Teilen zusammengesetzten Nachricht, so soll dieser gleich der Summe der Informationsgehalte dieser einzelnen Teilnachrichten sein. Die Wahrscheinlichkeit mit der die Gesamtnachricht erwartet wird ist aber - im Falle der statistischen Unabhängigkeit - gleich dem Produkt der Wahrscheinlichkeiten aller Teilnachrichten. Die Funktion f muss daher die Eigenschaft haben, dass die Summe mehrerer Funktionswerte gleich jenem Funktionswert ist, dessen Argument das Produkt der Argumente dieser einzelnen Funktionswerte ist, also

$$f(x) + f(y) = f(x*y)$$

Diese Forderung erfüllen nur die logarithmischen Funktionen:

$$\log(x) + \log(y) = \log(x*y)$$

Bleibt nur noch die Basis des Logarithmus offen. Wegen des Zusammenhanges zwischen dem Informationsgehalt einer Nachricht und ihrer Verschlüsselung durch einen Binärcode hat Shannon den Logarithmus zur Basis 2 (logarithmus dualis) für die Definition des Informationsgehaltes gewählt und die Einheit des Informationsgehaltes mit bit bezeichnet:

$$h = \mathrm{ld}(1/p) = - \mathrm{ld}\ p \ [\mathrm{bit}]$$

Die Bedeutung des Informationsgehaltes soll nun an einigen Beispielen illustriert werden:

Erhält ein Empfänger immer wieder die gleiche Nachricht, so kann die Wahrscheinlichkeit, mit der er diese Nachricht erwartet mit p=1 angenommen werden. Für diese sicher erwartete Nachricht ist jedoch der Informationsgehalt Null:

$$h = \mathrm{ld}\ 1 = 0 \ \mathrm{bit}$$

Eine solche Nachricht braucht erst gar nicht empfangen zu werden!

Eine völlig unerwartete Nachricht (p=0) hingegen hat - falls sie dennoch eintrifft - einen unendlich hohen Informationsgehalt.

Falls zwei alternative Nachrichten (zB ja oder nein) mit jeweils gleicher Wahrscheinlichkeit p=1/2 erwartet werden, so ist ihr Informationsgehalt 1bit:

$$h = \mathrm{ld}\ 2 = 1 \ \mathrm{bit}$$

Eine solche Nachricht liesse sich optimal durch ein einziges Bit codieren!

Werden Nachrichten mit unterschiedlicher Wahrscheinlichkeit erwartet, so ist der mittlere Informationsgehalt H einer solchen Nachricht gleich der mit den Wahrscheinlichkeiten p_i gewichteten Summe der Informationsgehalte h_i der einzelnen Nachrichten:

$$H = \sum p_i h_i = \sum p_i \mathrm{ld}(1/p_i) \ [\mathrm{bit}]$$

Falls n unterschiedliche Nachrichten mit gleicher Wahrscheinlichkeit p=1/n erwartet werden, so ist der mittlere Informationsgehalt einer solchen Nachricht

$$H = \mathrm{ld}\ n \ [\mathrm{bit}]$$

Der mittlere Informationsgehalt einer gleichwahrscheinlichen Nachricht, die durch eine Dezimalziffer dargestellt wird ist somit

$$H = ld\ 10 = 3.32 \text{ bit}$$

Für die binäre Codierung einer einzelnen Dezimalziffer müssen jedoch 4 bit verwendet werden. Umgekehrt können durch einen 4 Bit Code insgesamt 16 Zeichen verschlüsselt werden, 6 Codeworte bleiben somit ungenutzt. Dieser überflüssige Anteil wird als Redundanz bezeichnet. Die Redundanz R eines Codes ist die Differenz zwischen der mittleren Wortlänge L und dem mittleren Informationsgehalt H dieses Codes:

$$R = L - H \text{ [bit]}$$

Die mittlere Wortlänge L eines Codes variabler Länge ist die mit den Wahrscheinlichkeiten p_i gewichtete Summe der Wortlängen l_i der einzelnen Wörter dieses Codes:

$$L = \sum p_i l_i \text{ [bit]}$$

Auch die Redundanz wird in bit gemessen. Im Beispiel der Codierung gleichwahrscheinlicher Dezimalziffern durch einen 4 Bit Binärcode beträgt die Redundenz

$$R = 4 - 3.32 = 0.68 \text{ bit}$$

Um Nachrichten komprimiert speichern beziehungsweise rasch übertragen zu können ist eine geringe Redundanz wünschenswert. Ist die Redundanz Null, so ist der Code optimal komprimiert. In diesem Fall ist die mittlere Wortlänge gleich dem mittleren Informationsgehalt. Da die Redundanz nie negativ sein kann, ist der mittlere Informationsgehalt eine obere Schranke für die mittlere Wortlänge, das heisst die mittlere Wortlänge eines Codes kann nie kleiner sein als sein mittlerer Informationsgehalt. Es gilt immer

$$L \geq H$$

Die Redundanz eines Codes variabler Länge ist dann gering, wenn die Längen der einzelnen Codewörter möglichst gleich ihrem Informationsgehalt sind. Das bedeutet aber, dass häufig verwendete Codewörter kurz und selten verwendete lang sein müssen. Beispiele für dieses Prinzip sind die Kurzwahl für häufig verwendete Telefonnummern, Abkürzungen für häufig verwendete Bezeichnungen (zB Studi, Legi, Tram, PC) oder der Morsecode.

Andererseits bietet ein redundanzfreier Code keinerlei Möglichkeit der Fehlererkennung. Es genügt eine geringe Störung um zB aus Legi Vegi zu machen. Um solche Fehler erkennen oder gar korrigieren zu können bedarf es einer hohen Redundanz.

Die Bedeutung von Informationsgehalt und Redundanz kann am Beispiel der deutschen Sprache illustriert werden. Treten alle 26 Buchstaben mit gleicher Wahrscheinlichkeit in Texten auf, so ist der Informationsgehalt eines Buchstabens gleich ld 26 = 4.7 bit. In deutschen Texten treten diese Buchstaben jedoch mit unterschiedlicher Wahrscheinlichkeit auf. Mehr als die Hälfte aller Texte bestehen aus Leerzeichen und den Buchstaben e, n, r und i. Unter Berücksichtigung dieser Wahrscheinlichkeiten verringert sich der Informationsgehalt zu 4.1 bit pro Zeichen. Berücksichtigt man weiters, dass bestimmte Buchstabengruppen wie zB en, er, ch besonders häufig auftreten und verschlüsselt man diese gemeinsam, so sinkt der mittlere Informationsgehalt auf weniger als 2 bit pro Zeichen. Im Vergleich zur Verschlüsselung einzelner Buchstaben ergibt sich somit eine Redundanz von etwa 50%. Das bedeutet, dass im Mittel jeder zweite Buchstabe eines Textes weggelassen werden kann, ohne die Lesbarkeit zu beeinträchtigen.

zB D■E■■E■■NDA■ D■R■D■U■SCH■■■SPRA■H■ ■ET■AE■T ■T■■

F■■NFZ■G■P■■ZEN■

Einen ähnlichen Wert für die Redundanz erhält man, wenn man die Häufigkeit der Wörter der deutschen Sprache zur Berechnung des Informationsgehaltes heranzieht. Von den mehr als 10 Millionen Wörtern treten die drei häufigsten Wörter (die, der, und) mit einer Häufigkeit von 9.5% auf. Die 15 häufigsten Wörter machen ein Viertel aller deutschen Texte aus, während mit 66 Wörtern bereits die Hälfte abgedeckt wird. Auf Grund dieser Häufigkeitsverteilung berechnet sich der Informationsgehalt eines Wortes zu 11.8 bit. Da ein Wort im Mittel aus 5.7 Buchstaben besteht, erhält man einen mittleren Informationsgehalt von etwa 2 bit pro Buchstaben.

Völlig redundant - das heisst ohne Informationsgehalt - ist in der deutschen Sprache der Unterschied zwischen Gross- und Kleinschreibung, da dieser auf Grund grammatikalischer Regeln aus der Zeichenfolge rekonstruiert werden kann.

Die Stenographie macht von der Redundanz Gebrauch indem einzelne redundante Buchstaben (zB der Vokal e) völlig weggelassen werden und häufige Buchstabenfolgen durch Kürzel verschlüsselt werden.

Die hohe Redundanz natürlicher Sprachen macht die Fehlererkennung von Tippfehlern bei geschriebenen Texten oder auch das Verstehen gesprochener Texte trotz Nebengeräusche möglich.

Interessant ist auch die Frage, mit welcher Geschwindigkeit der Mensch Information aufnehmen kann. Beim Lesen erreicht der Mensch eine Geschwindigkeit von etwa 25 Buchstaben in der Sekunde, das entspricht einem Informationsfluss von 50 bit/s. Dieser Wert ist unabhängig von der verwendeten Sprache und dem Zeichenvorrat des Alphabets (der gleiche Informationsfluss kann auch beim Lesen von chinesischem Text erzielt werden). Auch akustische Nachrichten (zB gesprochener Text oder Musikdarbietungen) können mit einer Geschwindigkeit von maximal 50 bit/s wahrgenommen werden. Tatsächlich mit dem Bewusstsein verarbeitet wird davon höchstens die Hälfte, das sind 25 bit/s. Nimmt ein Mensch durch 60 Jahre hindurch täglich 16 Stunden lang Information mit dieser Geschwindigkeit auf, so erreicht er insgesamt einen Informationsgehalt von über $3*10^{10}$ bit. Theoretisch wäre der Mensch auch auf Grund der enormen Speicherkapazität des Gehirns von etwa 10^{12} bit auch in der Lage, diese Information zu speichern. Dennoch findet die gesamte im Laufe eines Lebens aufgenommene Information in optimal codierter Form auf einer CD Platz.

Für die Verschlüsselung der Erbinformation innerhalb der Gene verwendet die Natur einheitlich für alle Lebewesen einen aus 64 Zeichen bestehenden Genetischen Code. Als Gen bezeichnet man einen Abschnitt eines Kettenmoleküls der Desoxyribonukleinsäure (DNA) an dem in linearer Folge jeweils eine von vier möglichen Stickstoffbasen (Adenin, Uracil, Guanin, Cytosin) hängen. Je drei aufeinanderfolgende Basen bilden ein Wort. Insgesamt lassen sich $4^3 = 64$ verschiedene Worte bilden. Ein Gen enthält etwa 200 Worte, ein Chromosom enthält rund 10^4 bis 10^5 Gene. Die Anzahl der Chromosomen pro Zellkern ist für die einzelnen Lebewesen unterschiedlich und beträgt bei Wirbeltieren rund 50 (für den Menschen 46). Die pro Zellkern gespeicherten Daten haben somit ebenfalls ein Volumen von etwa 10^{10} bit.

Der Vergleich zeigt unter anderem die Mannigfaltigkeit auf die Daten in technischen und biologischen Systemen durch Signale dargestellt und übertragen werden können. Erst durch die individuelle Interpretation dieser Daten durch den Menschen entsteht Information. Werden Informationen miteinander verknüpft und Sachverhalten zugeordnet, so wird daraus Wissen und aus diesem gelegentlich Weisheit.